# Integrating Data Management Literacies with Data Visualization Instruction

## A One-Shot Workshop

LIBRARIES INDIANA UNIVERSITY Bloomington

Stacy Konkiel, Brianna Marshall & David Edward Polley | Indiana University-Bloomington

CNS Cyberinfrastructure for Network Science Center

## Introduction

This poster describes a drop-in data management and visualization workshop under development. The workshop, scheduled for October 2013, will be aimed at graduate student researchers but open to all.

The authors have mapped the core competencies for data information literacy outlined in "Determining Data Information Literacy Needs: A Study of Students and Research Faculty" (2011) with the workshop learning outcomes. Workshop planning considerations, content, and problems faced are described below. While this poster describes a workshop being developed at Indiana University, the authors hope this model will prove informative for other institutions looking to create similar workshops pairing data management and visualization.

## Planning Process

•Tools chosen are freely available; two are open source
•Team met weekly to discuss learning outcomes, data sources, data analysis techniques, and assessment strategies

Some ideas were discarded over the course of planning:

•Digital humanities data in HathiTrust (HT) and HT-related software were not publicly available for use, so the authors replaced HT data analysis with that of open humanities data provided by the IU Digital Collection Services. We plan to use the tool Voyant to visualize this data.

•"Frameworks for a Data Management Curriculum" (Lamar Soutter Library *et al*, 2012) course plans were originally intended to be adapted for this workshop. They proved to be too technical and detailed for incorporation into a beginner-level class.

## Download poster and handouts

http://hdl.handle.net/2022/16814

## DIL Core Competencies mapped to Learning Outcomes

| DIL Core Competencies | Learning Outcome | Skills Required | Assessment Measure |
|---|---|---|---|
| **Data Management and Organization**<br>**Introduction to Databases and Data Formats**<br>**Data Conversion and Interoperability** | ***Understand data management and organization concepts.*** | Identify data types, define data and its lifecycle, and describe how data is used and reused in research. List reasons why data management and organization is important to researchers. | Related questions on incoming and outgoing questionnaires. |
| **Data Visualization**<br>**Quality Assurance** | ***Design effective data visualizations.*** | Apply Illinsky's four principles of data visualization. Use Google Refine to clean data in order to assure data quality. Use Sci2 to perform various types of analysis on data sets. Use Gephi to visualize the data. | Critique visualizations according to rubric. |
| **Introduction to Databases and Data Formats**<br>**Discovery and Acquisition of Data**<br>**Data Analysis** | ***Find and select appropriate data—and apply proper analysis—to create a visualization that answers a particular research question.*** | Identify research question. Choose appropriate types and formats of data for topical, network, burst, and temporal analysis. Navigate to data sources. Download data in proper format. Analyze data. | Successful completion of questionnaire addressing skills needed to answer research question. |
| **Cultures of Practice**<br>**Introduction to Databases and Data Formats** | ***Understand how cultures of practice influence the way data may be collected, described, or formatted.*** | Based on intended audience for visualization and source of data, identify culture of practice/discipline. Be able to access information on discipline's data collection standards [literature search], relevant metadata schema and controlled vocabularies [libraries website and DCC guide], and what tools and formats are common to particular disciplines [literature search]. | Successful literature searches (and reading) for data collection and analysis methodologies. Successful access of disciplinary metadata schema. |
| **Metadata**<br>**Data Preservation** | ***Save data to IU-supported research storage for both short- and long-term storage.*** | Identify best storage options for short- and long-term data storage. Successfully navigate to appropriate storage solutions, log in. Transfer data from local drives to cloud storage accounts. | Questionnaire. |
| **Ethics (including citation of data)**<br>**Data Preservation** | ***Handle data and data visualizations ethically.*** | Cite data from other sources in visualizations and documentation. Decide which data to make available, based on sensitivity. Store data on appropriate technologies using safeguards, based on sensitivity. Create visualizations that accurately represent the source dataset (i.e. does not manipulate or skew results). | Questionnaire.<br>[In a more in-depth workshop, one idea would be to have students create projects which are then reviewed and scored against a rubric after class ends.] |
| **Data Curation and Re-use**<br>**Ethics (including citation of data)**<br>**Data Management and Organization** | ***Prepare data and visualizations for re-use.*** | Create documentation for others to reference when reusing data that describes methodology for finding, analyzing, and visualizing data. Cite data source and visualization creator in visualization caption(s). Organize and format data appropriately for future processing by tools. | Questionnaire.<br>[In a more in-depth workshop, one idea would be to have students create projects which are then reviewed and scored against a rubric after class ends.] |

## Considerations

### Designing our tutorials

•Emphasize analysis over datasets themselves
•Future workshops responsive to feedback

### Finding appropriate data sources

•Diversity of available datasets
•Chose a wide breadth of examples to fit our expected audience
•Criteria for selection: usable format, properly documented, serendipity

### Assessing the learning outcomes

•Difficult in one-off workshops
•Self-reporting of understanding less than perfect
•Project-based assessment would be ideal

## References

Carlson, Jake R.; Fosmire, Michael; Miller, Chris; and Sapp Nelson, Megan, "Determining Data Information Literacy Needs: A Study of Students and Research Faculty" (2011). *Libraries Faculty and Staff Scholarship and Research*. Paper 23. http://docs.lib.purdue.edu/lib_fsdocs/23

Lamar Soutter Library, University of Massachusetts Medical School and the George C. Gordon Library, and Worcester Polytechnic Institute. (2012). "Frameworks for a Data Management Curriculum: Course plans for data management instruction to undergraduate and graduate students in science, health sciences, and engineering programs." Retrieved from http://library.umassmed.edu/data_management_frameworks.pdf

## Contact

skonkiel@Indiana.edu // @skonkiel

bhmarsha@umail.iu.edu // @nososternlib

## Workshop Specifics

**Analyses**

- Topical
- Temporal
- Spatial
- Network

**Data Sources**

- ISI / Web of Knowledge
- Open Congress
- ICPSR
- Open humanities data from IU

**Tools**

Sci2 A tool for science of science research & practice

Gephi

Google refine

VOYANT